Python Machine Learning Intrusion Detection System Proposal

Collin Pike Jordan Sosnowski John David Watts

2020-3-1

Modern networks are constantly under attack from malicious agents, such as malicious insiders, advanced persistent threats, nation-state actors, hacktivists, etc. If a network is breached it can cost businesses hundreds of millions of dollars; therefore, network security is of utmost importance. Unfortunately, breaches can have more than just economic repercussions. Employees' data can be leaked and their integrities compromised, which can lead to the loss in trust of the affected company. To protect against these attacks, it is imperative to have a network that is up to date and analyze network traffic for attacks. However, manually analyzing data streams is feasibly impossible, especially for large networks. To combat this, intrusion detection systems (IDS) can be used to slim down the amount of data analysts have to sift through.

Unfortunately, most IDSs on the network today are heuristic-based; by that we mean the framework analyzes packets against known signatures. If a packet, or stream of packets, match these signatures the system will raise an alert. This would be a perfect solution if network attacks were stagnant and did not continue to evolve, however, in this world attacks are constantly changing and evolving. Since we have this cat and mouse game with attackers signature-based heuristics will always be a step behind, as signatures as based on known attacks.

Therefore, we propose a different type of IDS. A machine learning anomaly-based intrusion detection system. One that determines it's networks 'normality'. Once 'normality' is achieved anything that is not within the realms of normal can be assumed to be a threat.

To keep this in the scope of a semester project we want to see if our machine learning framework can detect two different types of attacks: a syn flood, and a Nmap port scan. If we finish with time left we would like to scope it out to see if it can detect other types of attacks. Since it is not heuristic-based, in theory, it should be able to detect many attacks that we have not thought of.